DOCUMENT RESUME

ED 359 195 TM 019 815

AUTHOR	Sutton, Rosemary E.
TITLE	Equity Issues in High Stakes Computerized Testing.
PUB DATE	29 Mar 93
NOTE	16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 11-16, 1993).
PUB TYPE	Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Access to Education; *Adaptive Testing; *Computer Assisted Testing; Educational History; Elementary Secondary Education; *Equal Education; Test Bias; Testing Problems; *Test Use; Test Wiseness; Trend Analysis
IDENTIFIERS	*High Stakes Tests

ABSTRACT

This paper discusses equity issues that may arise from the widespread use of high stakes computerized testing. The literature relevant to computerized testing is examined from the perspective of equity concerns from within the framework of research and from the perspective of possible uses of computerized testing if equity issues are considered paramount. In the first instance, the question concerns whether the use of computerized testing will maintain or exaggerate inequities in education. In the second section, the question is how the new technology can be used to reduce inequities. The following six topics are considered in examining whether computerized testing will maintain or exaggerate inequities: (1) equivalence; (2) prior experience; (3) setting of computers; (4) long-term attitudes toward computerized testing; (5) expectancies and adaptive testing; and (6) testwiseness and adaptive testing. Issues in the equity advocate approach (the role of computerized testing in reducing inequity) include time factors, guessing and adaptive tests, and the format and style of questions. The history of computer use in schools and testing suggests that inequities will be maintained or exaggerated, but it is possible to use computerized testing to reduce inequities if that is a recognized goal. (Contains 58 references). (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED359195

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
ROSEMARY E. SUTTON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Equity Issues in High Stakes Computerized Testing

Rosemary E. Sutton

College of Education
Cleveland State University
Cleveland, OH 44115
(216) 687-4573
(216) 687-9366(FAX)

March 29, 1993

Paper presented at the 1993 Annual Meeting of the American Educational Research Association, Atlanta, GA.

TM019815

BEST COPY AVAILABLE

Equity Issues in High Stakes Computerized Testing

While computerized testing has been developed and used for more than a decade, recent advances in technology and item response theory mean its use is about to become widespread (Wise & Plake, 1990). By Fall, 1993, ETS plans to have computerized versions of the SAT (Scholastic Achievement Test), and GRE (Graduate Record Examination) in use. These initial computerized tests will be similar in nature to pen and paper versions in that many questions will be presented in a linear fashion, and students will be able to skip items and go back to them later if they wish. The computerized versions are different from the traditional versions in that they make use of the computer capabilities of graphics and movement (the interface Windows is being used), and a mouse and keyboard are used for input. Later versions of the tests will be adaptive. This means that the initial items in any test will identify a student's level of competence. Once this has been established the program branches to items that are appropriate for that student. A total test length of approximately 20 items is sufficient to identify students' performance with satisfactory reliability. These adaptive tests are much shorter than traditional tests, but will not allow students to skip items and go back as later items depend on the performance on earlier items.

A large research literature has documented inequities in computer uses in education in K-12 (Sutton, 1992) and tertiary education (e.g. Dambrot, Watkins-Malek, Silling, Marshall & Garver, 1985; Temple and Lips, 1989). The use of microcomputers in schools during the 1980's maintained and exaggerated existing inequities in education. The focus of this paper is on equity issues that may arise from the widespread use of high stakes computerized testing. I examine the literature relevant to computerized testing from two perspectives. First, equity concerns from within the framework of research on testing are considered. In this section, labeled benevolent psychometric, the essential question is - will the use of computerized testing maintain or exaggerate inequities in education? Second, the possible uses of computerized testing if equity issues are considered paramount are discussed. In this approach, labeled equity advocate, the question is - how can this new technology be used to reduce inequities?

Benevolent Psychometric Approach

In this approach equity issues are considered within the framework of psychometric theory and practice. The concern is whether the use of computerized testing maintains or exaggerates inequities in education. Six topics are relevant to this question: the concept of equivalence of scores, the role of prior experience, the setting (public vs. private) of computers, long term attitudes, testwiseness, and expectancies in computer adaptive testing.

Equivalence.

A typical paradigm for research and development on computerized testing is to compare the scores on two versions of the test in order to establish equivalence.

> "Scores from conventional and computer administrations may be considered equivalent when (a) rank orders of scores tested in alternative modes closely approximate each other, and (b) means, dispersions, and shapes of the score distributions are approximately the same, or have been made approximately the same by rescaling the scores from the computer mode (APA, 1986, pp. 13-14).

This approach makes two assumptions relevant to equity. First, it focuses on group rather than individual differences. In most of the studies on equivalence scores on the computerized tests have been lower than on the conventional tests, but typically these differences have not been statistically significant, and thus have been nonsignificant and considered too small to be meaningful (Bunderson, Inouye, & Olsen, 1989). These differences, however, may be due to a substantially poorer performance of a small proportion of examinees (Wise, Barnes, Harvey, & Plake, 1989) and research is needed to examine this.

Second, this approach assumes that the status quo is an acceptable baseline. Inequities that may exist in conventional testing, are not considered relevant. For example Green, Bock, Humphreys, Linn and Reckase (1984) stated that "for equity, equivalence of expected scores on the two forms is sufficient" (p. 357). Evidence does exist, however, that there are inequities in existing standardized testing. For example, the SAT-quantitative scale has consistently shown male superiority (in 1992, males on average scored 43

points higher) even though few gender differences in computation and mathematical concepts exist for the general population when assessed by other measures (Kimble, 1989; Hyde, Fennema & Lammon, 1989; Linn, & Hyde, 1989). Many people believe have argued that existing standardized tests are biased against poor and minority students (e.g., Hacker, 1992) and low achieving students (e.g., Paris, Lawton, Turner & Roth, 1991); if computerized tests replicate these patterns these inequities are maintained.

Prior Experience.

Experience in using computers has been found to be related to computer-related competence (Martinez & Mead, 1988) and attitudes towards computers (e.g., Arenz & Lee, 1990; Gressard & Loyd, 1987). Data from the early and mid 1980's clearly demonstrated that poor and minority children had less access to computers at home and at school (e.g., Becker, 1983, Oct; Becker & Sterling, 1987; Martinez & Mead, 1988). Female students had less access at home (e.g., Arenz & Lee, 1990, Chen 1986), and in many schools (Becker & Sterling, 1987). College students at more selective colleges were more likely to own and use computers than college students at less selective colleges (Turner, 1987). Older students have also been found to have less knowledge about computers (Massoud, 1991) and older adults have been reported to have less favorable attitudes towards computers (Baack, Brown & Brown, 1991; Morris, 1988-89).

This lower access and lesser experience could impact on students' performance in computerized testing. While the tests begin with a tutorial on the use of the computer and mouse, such a tutorial cannot provide enough experience to have computer use become automatic. Some research has documented that prior computer experiences are significantly related to computerized testing performance (e.g., Johnson & White, 1980; Lee, 1986) whereas other research has found no negative effects for lack of experience (Wise, Boettcher-Barnes, Harvey & Plake, 1989). These findings may be very dependent on the kind of tutorial given at the beginning of the test, the range of experience of the subjects, and the design and complexity of the software. It is important to study whether differential experience affects test performance for each version of the test.

Setting of Computers.

Several studies conducted at Princeton University have demonstrated that the setting (private vs. public) of a computer influences attitudes and anxiety. Cooper, Hall & Huff (1990) reported that both boys and girls liked a "male" computer program (Demolition Division) more than the "female " program (American Classroom fractions) but that the stress was higher for boys and girls when using the cross-gender program in a public context (computer center) compared to working in privacy. Robinson-Staveley and Cooper (1990) had female and male college undergraduates complete a difficult computer task and a series of questionnaires in the presence of absence of another person. For women, with little computer experience, those who worked in the presence of another performed less well and reported more anxiety than did women who worked alone. For low-experienced men mere presence had the opposite effect. The setting did not affect the performance and attitudes for high experienced men and women. A follow-up experiment manipulated expectancies for success and found that the presence of another person hindered low expectancy students but facilitated high expectancy students. The presence of another person consisted of a individual working on a separate computer, facing a different wall from that of the subject, and making no verbal interaction.

It is important to determine if these results apply to computerized testing. The testing centers which ETS is setting up will contain a number of computers perhaps with some kind of petition separating computers and users. Will this constitute a "public setting" and influence the performance of low expectancy and female students? It is possible that this same affect now occurs with the use pen and paper testing but it has not been studied empirically. One of the aspects of working on computers that is different from pen and paper is the relatively public nature of monitor screens so it is possible that this affect is unique to computer use.

Long Term Attitudes towards Computerized Testing.

Studies have indicated that more experience with computers is related to more positive attitudes

for students attending elementary school (Lever, Sherrod, Bransford, 1989), junior high and high school (Arenz & Lee, 1990; Loyd & Loyd, 1988), and college (Loyd & Loyd, 1988; Wu & Morgan, 1989). Of course, students have attitudes towards standardized testing as well as computer use. Paris et al. (1991) have documented that these attitudes decline with age: older students were more suspicious about test validity, reported decreasing motivation to excel, and felt less prepared to take tests. It is important to study attitudes towards high stakes computerized testing as these tests are introduced and used.

The results of existing research on attitudes towards computerized achievement testing are conflicting. Several studies using volunteers have reported very positive attitudes even by students with little prior experience (Schmidt, Urry & Gugel, 1978; O'Neill & Kubiak, 1992; Ward, 1988) and this includes two studies using software developed by ETS (O'Neill & Kubiak, 1992; Ward, 1988). However, there were no negative (or positive) consequences associated with scores gained on these tests, and the novelty effect may have been strong (Ward, 1988). In contrast, two studies in which the computerized tests counted for college course grade, reported negative attitudes and higher anxiety (Gwinn & Beal, 1987-88; Ward, Hooper & Hannafin, 1989) although in one of these studies the majority of students reported preferring computerized testing to pencil and paper tests (Gwinn & Beal, 1987-88). It is impossible to predict the range of attitudes that may develop towards high stakes computerized testing. Only by continuously assessing attitudes during implementation and use will we begin to understand students' feelings and beliefs about this new use of technology, and what equity issues develop.

Expectancies and Adaptive Testing.

In traditional achievement test theory (Gronlund, 1971; Nunnally, 1964) item difficulties of approximately .50 are sought because this leads to high levels of discrimination amongst the testees. This is, however, as average item difficulty for the population of test takers and typically results in less discrimination for very high and very low achieving students. In computer adaptive testing the difficulty levels of the later items in the test are tailored to the individual student so that for each student the difficulty

level is approximately .50. For very high achieving students, used to performing very well on tests, these items will seem much harder than those they are accustomed to. For low achieving students, the reverse is true: their items will also be at the 50% level of difficultly (for them) and will be easier than what they are accustomed to.

It is not clear how this will affect attitudes, expectancies and performance of students taking these adaptive tests. Research on motivation has distinguished between students and situations that are performance (or ego) oriented where the goal is to seek positive judgements and avoid negative judgements of competence from those that are learning (or task) oriented where the goal is to increase competence (e.g. Dweck, 1986). In high stakes testing situations performance goals are very salient. Research has shown that students with low assessments of their own ability in such conditions often choose personally easy tasks in which their success is assured or excessively difficult ones on which their failure does not signify low ability (Dweck, 1986). How will low achieving students with low assessments of their own abilities react to adaptive tests where the level of difficulty for them is approximately .50? How about high achieving students with low assessments of their own ability? (this group appears to be disproportionately female) Might some students figure out that one strategy is to answer initial items wrong so that easier items are presented? Only thorough and long term research programs will answer these questions.

Testwiseness And Adaptive testing

Testwise examinees attain improved scores on tests by using test-taking strategies that are construct-irrelevant but make the test easier for them (Messick, 1989). Four types of strategies test-wise examinees understand and use are time-using strategies, error-avoidance strategies, guessing strategies, and deductive reasoning strategies (Millman, Bishop & Ebel, 1965; Sarnacki, 1979). Research has shown that testwise students do score higher (e.g., Rogers & Bateson, 1991) and that these strategies can be taught and lead to higher test scores among white, Black and Hispanic students (Dreisbach & Keogh, 1982; Kalechstein, Kalechstein & Docter, 1981; Maspons & Llabre, 1985; Sarnacki, 1979).

Under computerized adaptive testing testwise strategies may vary. For example, students are

typically taught to skip or omit items that appear difficult and return to them later. Under adaptive testing items cannot be skipped. Testwise students typically work quickly through a test and check all items if time remains. In adaptive testing item checking is not possible. There is no way to determine whether the changes in test taking strategies that adaptive testing will demand will help or hinder those groups typically not testwise. The coaching academies for the college entrance tests would be expected to continue (with some modification of content) and thus continue to advantage higher income students.

Equity Advocate Approach

In this approach equity issues as perceived as paramount rather than traditional psychometric concerns. The question to be examined is - can computerized testing be used to reduce inequities in testing and education? Areas to be discussed in this section are time to take computerized tests, guessing on multiple choice items, and the variety of formats and items computerized testing allow.

I am not assuming in this approach that there should be no differences among individuals in test scores. It also may be possible that there are genuine social class (and therefore ethnic) score differences in test taking by the college level as there may accumulated cognitive deficits as a result of poorer schooling and poverty (Ginsburg, 1986). However, I do believe that if such differences in mean group test scores can be altered by the time allotted to take the test, minimizing the impact of guessing, or altering the test format then these do not represent cognitive deficits but are artifacts of testing conventions.

Time to Take the Test.

A consistent finding in research on computerized linear achievement testing is that it takes less time than conventional testing; computer adaptive testing requires many fewer items than linear testing to establish reliability so the tests are much shorter and thus take even less time (Bunderson, Inouye & Olsen, 1989; Olsen, 1990) This reduction in time has been seen as an increase in efficiency and as allowing more

time for students to be engaged in instructional activities (Wise & Plake, 1989). However, from an equity perspective, this "increase in efficiency" should be used to allow students more time to complete test items.

Research has demonstrated that unlimited time to take tests helps narrow the gap between female and male SAT quantitative tests (Dreyden & Gallagher, 1989; Gallagher & Johnson, 1992) and that Black, Mexican American, and Puerto Rican students take more time to complete tests (Llabre, 1991; Llabre & Froman, 1985; Schmitt & Dorans, 1990). Experimental studies allowing Black and Hispanic examinees more time have frequently shown that their performance relative to whites is not enhanced (Evans & Reilly, 1972, 1973; Wild, Durso & Rubin, 1982), but there are methodological shortcomings in the research (Llabre, 1991) and these studies have not allowed unlimited time. Unlimited time is conceptually very different for a test taker than "more" time, and the recent research on female performance in the quantitative SAT suggests that this is a fruitful area of research. From an equity perspective, the goal should be to determine how much the gap between groups is narrowed by allowing unlimited or longer periods of time to take computerized tests. [1]

Guessing and Adaptive Tests

In multiple choice tests if there is no penalty for incorrect answers, students should guess even when they do not know the correct answer. If a penalty for incorrect answers exists a conservative guessing strategy will tend to lead to an increased test score (Slakter, 1968). Females tend to omit more items and guess less in multiple choice tests (Ben-Shakhar & Sinai, 1991; Slakter, Koehler, Hampton, Grennell, 1971). In computer adaptive testing omitting items is not possible because continuation to the next item depends on completion of the prior item. This could benefit females although the tendency to guess less seems to account for a small fraction of the gender differences in achievement.

---

[1] This suggestion, of course, is counter to long standing western assumptions that speededness is an important component of intelligence and achievemnent.

Format and Style of Questions

Research from Great Britain suggests that females perform less well on multiple choice tests than more open ended formats (Bolger & Kellaghan, 1990; Murphy, 1982). Computerized testing allows for more open ended formats under two different scoring techniques. First, key words can be identified in examinees' responses by the computer program and so scoring is immediate. This is most appropriate when specific words or phrases are sought, e.g. a synonym or a name of a country. In addition, more open ended constructed responses can be included and the responses transmitted on line to be scored at some agency (e.g. ETS). An example of this type of item might be to ask examinees to read a passage containing contradictions and ask them to generate possible hypotheses for these contradictions. Scoring by humans is obviously much more time consuming, and therefore costly, and means that students' cannot receive immediate total test scores.

Whether more open ended formats in computerized testing will reduce the female SAT disadvantage is not clear. The hypotheses for females' poorer performance on multiple choice tests are varied. Some argue that the disadvantage on multiple choice tests results from females' greater verbal skills (Murphy, 1982). It has also been proposed that females have neater handwriting which influences examination scoring (Murphy 1982) but under computerized testing this is obviously not relevant. A common explanation is female's tendency to omit more items because of lower rates of guessing and risk taking, and this was discussed above. It has also been suggested that females have different ways of knowing (Belenky, Clinchy, Goldberger & Tarule, 1986) and that their knowledge is more contexualized. If this is true multiple choice tests are a poor match for this kind of knowledge.

The computer capabilities of color graphics and animation allow for items of more variety than traditional pen and paper tests. Boykin (1978) argued that because of the high intensity and variability of home and immediate ecological environments, African American children find tasks presented in a relatively monotonous fashion even more intolerable than their White counterparts. A study using 3rd grade students found that high task variability resulted in significantly higher scores for African American students, but did not affect the performance of White students (Boykin, 1982). It is unknown whether these findings would

apply to the testing conditions of African American students in the 1990's but from an equity point of view differential effects for format variability are worth exploring.

Conclusions

How we proceed with computerized testing will reflect our values as an education community and society. Will we replicate existing inequities and select the conventional approach to equivalence in computerized testing, or will we actively seek to use this technology to help female students and students of color? The history of computer use in schools and testing suggests inequities will be maintained or exaggerated. The choice, however, is ours.

References

American Psychological Association Committee on Professional Standards (COPS) and Committee on Psychological Tests and Assessment (CPTA). (1986). Guidelines for computer-based tests and interpretations, Washington DC: Author.

Arenz, B. W., & Lee, M. J. (1990, April). Gender differences in the attitudes, interest and participation of secondary students in computer use. Paper presented at the Annual meeting of the American Educational Research Association, Boston.

Baack, S. S., Brown, T. S., Brown, J. T. (1991). Attitudes towards computers: Views of older adults compared with those of young adults. Journal of Research on Computing in Education, 23, 422-433.

Becker, H. J. (1983, October). School uses of computers: Reports from a National Survey, No. 3. Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Becker, H. J. & Sterling, C. W. (1987). Equity in school computer use. National data and neglected considerations. Journal of Educational Computing Research, 3, 289-311.

Belenky, M., F., Clinchy, B. M., Goldberger, N. R., Tarule, J. M. (1986). Women's ways of knowing: The development of self, voice and mind, New York: Basic Books.

Ben-Shakhar, G., & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. Journal of Educational Measurement, 28 (1), 23-35.

Boykin, A. W. (1978). Psychological/behavioral verve in academic task performance: pre-theoretical considerations. Journal of Negro Education, 47, 343-354.

Boykin, A. W. (1982). Task variability and the performance of black and white school children: Vervistic explorations. Journal of Black Studies, 12, 469-485.

Bunderson, C. V., Inouye, D. K. & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), Educational Measurement, 3rd ed.(pp. 367-408). New York: Macmillan.

Chen, M. (1986). Gender and computers.: The beneficial effects of experience on attitudes. Journal of Educational Computing Research, 2, 265-282.

Cooper, J. Hall, J., & Huff, C. (1990). Situational stress as a consequence of sex-stereotyped software. Personality and Social Psychology Bulletin, 16, 419-429.

Dambrot, F. H., Watkins-Malek, M. A., Silling, S. M., Marshall, R. S., Garver, J. A. (1985). Correlates of sex differences in attitudes toward and involvement with computers. Journal of Vocational Behavior, 27, 71-86.

Deisbach, M., & Keogh, B. K. (1982). Testwiseness as a factor in readiness test performance of young Mexican-American children, Journal of Educational Psychology, 74, 224-229.

Dreyden, J. I., & Gallagher, S. A. (1989). The effects of time and direction on the SAT performance of academically gifted adolescents. Journal for the Education of the Gifted, 12,(3) 187-204.

Dweck, C. S. (1986). Motivational processes affecting learning. American Psychologist, 41, 1040-1048.

Evans, F. R. & Reilly, R. R. (1972). A study of speededness as a source of test bias. Journal of Educational Measurement, 9 (2), 123-131.

Evans, F. R., & Reilly, R. R. (1973). A study of test speededness as a potential source of bias in the quantitative score of the admission test for graduate study in business. Research in Higher Education, 1, 173-183.

ETS, (1992) Development of a computer-based test for the GRE general test. Symposium presented at the 1992 Annual Meeting of the National Council on Measurement in Education, San Francisco, California.

Gallagher, S. A., & Johnson, E. S. (1992). The effect of time limits on performance of mental rotations by gifted adolescents. Gifted Child Quarterly, 36 (1), 19-26.

Ginsburg, H. P. (1986). The myth of the deprived child: New thoughts on poor children. In U. Neisser, (Ed), The school achievement of minority children: New perspectives (pp. 169-189). Hillsdale, NJ:, Erlbaum.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. Journal of Educational Measurement, 21, 347-360.

Gressard, C. P. & Loyd, B. H. (1987). An investigation of the effects of math anxiety and sex of computer attitudes. School science and mathematics, 87 (2), 125-135.

Gronlund, N. E. (1971). Measurement and evaluation in teaching, (2nd ed)., New York: Macmillan.

Gwinn, J. F., & Beal, L. F. (1987-88). On-line computer testing: Implementation and endorsement. Journal of Educational Technology Systems, 16, 239-252.

Hacker, A. (1992). Two nations: Black and white, separate, hostile, unequal. New York: Scribners.

Hyde, J. S. Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analyses. Psychological Bulletin, 107, 139-155.

Johnson, D. F., & White, C. B. (1980). Effects of training on computerized test performance in the elderly. Journal of Applied Psychology, 65, 357-358.

Kalechstein, P., Kalechstein, M., & Docter, R (1981). The effects of instruction on test-taking skills in second-grade black children. Measurement and Evaluation in Guidance, 13 (4), 198-202.

Lee, M. W. (1986). The match: Learning styles of black children and microcomputer programming. Journal of Negro Education, 55 (1), 78-90.

Lever, S., Sherrod, K. B., & Bransford, J. (1989). The effects of Logo instruction on elementary students' attitudes towards computers and school. Computers in the Schools, 6 (1/2), 45-65.

Linn, M. C., & Hyde, J. S. (1989, November). Gender, mathematics and science. Educational Researcher, 18, 17-19, 22-26.

Llabre, M. M. (1991). Time as a factor in the cognitive test performance of Latino college students. In G. D. Keller, J. R. Deneer & R. J. Magallan (Eds), Assessment and access: Hispanics in higher education (pp. 95-104), Albany, NY: SUNY.

Llabre, M.M., & Froman, T. W. (1986). Allocation of time to test items: A study of ethnic differences. Journal of Experimental Education, 137-140.

Loyd, B. H., & Loyd, D. E. (1988). Computer attitudes: Differences by gender and amount of computer experience. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans,

Martinez, M. E., & Mead, N. A. (1988). Computer competence: The first national assessment (Tech rep. No. 17-CC-01). Princeton, NJ: National Assessment of Educational Progress & Educational Testing Service.

Maspons, M. M., & Llabre, M. M. (1985). The influence of training Hispanics in test taking on the psychometric properties of a test. Journal for Research in Mathematics Education, 16, 177-183.

Massoud, S. L. (1991). Computer attitudes and computer knowledge of adult students. Journal of Educational Computing Research, 7 269-291.

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-726.

Morris, D. C. (1988-89). A survey of age and attitudes towards computers. Journal of Educational Technology Systems, 17 (1), 73-78.

Nunnally, J. C. (1967,1974). Educational measurement and evaluation, New York: McGraw-Hill,

Olsen, J. B. (1990). Applying computerized adaptive testing in schools. Measurement and evaluation in counseling and development, 23 (1) 31-38.

Robinson-Staveley, K., & Cooper, J. (1990). Mere presence, gender, and reactions to computers: Studying human-computer interaction in the social context. Journal of Experimental Social Psychology, 26 168-183.

Rogers, W. T., & Bateson, D. J. (1991). The influence of test-wiseness on performance of high school seniors on school leaving examinations. Applied Measurement in Education, 4, 159-183.

Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991, June-July). A developmental perspective on standardized achievement testing. Educational Researcher, 20, 12-20.

Sarnacki, R. E., (1979). An examination of test-wiseness in the cognitive domain. Review of Educational Research, 49, 252-279.

Schmitt, A. P., & Dorans, N. J. (1990). Differential item functioning for minority examinees on the SAT. Journal of Educational Measurement, 27, (66-81).

Slakter, M. J., Koehler, R. A., Hampton, S. H., Grennell, R. L. (1971). Sex, grade level, and risk taking on objective examinations. The Journal of Experimental Education, 39 (3), 65-68.

Sutton, R. E. (1991). Equity and computers in the schools: A decade of research. American Educational Research Journal, 61, 475-503.

Temple, L. & Lips, H. M. (1989). Gender differences and similarities in attitudes towards computers. Computers in Human Behavior, 5, 215-226.

Turner, J. A. (1987, April 29). Microcomputers found more available - and more often required - at selective colleges. Chronicle of Higher Education, 13,17.

Ward, T. J., Hooper, S. R., & Hannafin, K. M. (1989). The effects of computerized tests on the performance and attitudes of college students. Journal of Educational Computing Research, 5, 327-333.

Wild, C. L., Durso, R., & Rubin, D. B. (1982). Effect of increased test-taking time on test scores by ethnic, years out of school, and sex. Journal of Educational Measurement, 19 (1), 19-28.

Wise, S. L., Boettcher-Barnes, L., Harvey, A. L., & Plake, B. S. (1989). Effects of computer anxiety and computer experience on the computer-based achievement test performance of college students. Applied Measurement in Education, 2, 235-241.

Wise, S. L., Barnes, L. B., Harvey, A., & Plake, B. S. (1989). The effects of computer anxiety and computer experience on the computer-based achievement test performance of college students. Applied Educational Measurement in Education, 2, 235-241.

Wise, S. L., & Plake, B. S. (1989). Research on the effects of administering tests via computers. Educational Measurement: Issues and Practice, 8 (3), 5-10.

Wise, S. L., & Plake, B. S. (1990). Computer-based testing in higher education. Measurement and evaluation in counseling and development, 23 (1), 3-10.

Wu, Y-K. & Morgan, M. (1989). Computer use, computer attitudes and gender: differential implications of micro and mainframe usage among college students. Journal of Research on Computing in Education, 21, 214-228.